Regularized LRT for Large Scale Covariance Matrices: One Sample Problem

Young-Geun Choi, Chi Tim Ng and Johan Lim*

June 13, 2016

Abstract

The main theme of this paper is a modification of the likelihood ratio test (LRT) for testing high dimensional covariance matrix. Recently, the correct asymptotic distribution of the LRT for a large-dimensional case (the case p/n approaches to a constant $\gamma \in (0, 1]$) is specified by researchers. The correct procedure is named as corrected LRT. Despite of its correction, the corrected LRT is a function of sample eigenvalues that are suffered from redundant variability from high dimensionality and, subsequently, still does not have full power in differentiating hypotheses on the covariance matrix. In this paper, motivated by the successes of a linearly shrunken covariance matrix estimator (simply shrinkage estimator) in various applications, we propose a regularized LRT that uses, in defining the LRT, the shrinkage estimator instead of the sample covariance matrix. We compute the asymptotic distribution of the regularized LRT, when the true covariance matrix is the identity matrix and a spiked covariance matrix. The obtained asymptotic results have applications in testing various hypotheses on the covariance matrix. Here, we apply them to testing the identity of the true covariance matrix, which is a long standing problem in the literature, and show that the regularized LRT outperforms the corrected LRT, which is its non-regularized counterpart. In addition, we compare the power of the regularized LRT to those of recent non-likelihood based procedures.

Keywords: Asymptotic normality; covariance matrix estimator; identity covariance matrix; high dimensional data; linear shrinkage estimator; linear spectral statistics; random matrix theory; regularized likelihood ratio test; spiked covariance matrix.

1 Introduction

High dimensional data are now prevalent everywhere that include genomic data in biology,

*Young-Geun Choi and Johan Lim are at Department of Statistics, Seoul National University, Seoul, Korea. Chi Tim Ng is at Department of Statistics, Chonnam National University, Gwangju, Korea. All correspondences are to `johanlim@snu.ac.kr`

financial times series data in economics, and natural language processing data in machine learning and marketing. The traditional procedures that assume that sample size n is large and dimension p is fixed are not valid anymore for the analysis of high dimensional data. A significant amount of research are made to resolve the difficulty from the dimensionality of the data.

This paper considers the inference problem of large scale covariance matrix whose dimension p is large compared to the sample size n. To be specific, we are interested in testing whether the covariance matrix equals to a given matrix; $\mathcal{H}_0 : \Sigma = \Sigma_0$, where Σ_0 can be set I_p without loss of generality. The likelihood ratio test (LRT) statistic for testing $\mathcal{H}_0 : \Sigma = \mathrm{I}_p$ is defined by

$$\mathrm{LRT} \quad = \quad \mathrm{tr}\big(\mathbf{S}_n\big) - \log|\mathbf{S}_n| - p = \sum_{i=1}^{p} \big(l_i - \log l_i - 1\big),$$

where $\mathbf{S}_n$ is the unbiased and centered sample covariance matrix and l_i is the $i-$th largest eigenvalue of the sample covariance matrix. When p is finite, LRT follows the chi-square distribution with degrees of freedom $p(p+1)/2$ asymptotically. However, this does not hold when p increases. Its correct asymptotic distribution is computed by Bai et al. (2009) for the case p/n approaches $\gamma \in (0,1)$ and both n and p increase. They further numerically show that their asymptotic normal distribution defines a valid procedure for testing $\mathcal{H}_0 : \Sigma = \mathrm{I}_p$. The results of Bai et al. (2009) are refined by Jiang et al. (2012), which include the asymptotic null distribution for the case $\gamma = 1$. Despite of the correction of the null distribution, the sample covariance is known to have redundant variability when p is large, and it still remains a general question that the LRT is asymptotically optimal for testing problem in the n, p large scheme.

In this paper, it is shown that the corrected LRT can be further improved by introducing a linear shrinkage component. In detail, we consider a modification of the LRT, denoted by regularized LRT (rLRT), defined by

$$\mathrm{rLRT} = \mathrm{tr}\big(\widehat{\Sigma}\big) - \log|\widehat{\Sigma}| - p = \sum_{i=1}^{p} \big(\psi_i - \log\psi_i - 1\big), \tag{1}$$

where $\widehat{\Sigma}$ is a regularized covariance matrix and ψ_i is the $i-$th largest eigenvalue of $\widehat{\Sigma}$. Here, we consider the regularization via linear shrinkage:

$$\widehat{\Sigma} \equiv \lambda\mathbf{S}_n + (1-\lambda)\mathrm{I}_p. \tag{2}$$

We also occasionally notate $\mathrm{rLRT}(\lambda)$ to emphasize the use of the value λ. The linearly shrunken sample covariance matrix (simply shrinkage estimator) is known to reduce expected estimation loss of the sample covariance matrix (Ledoit and Wolf, 2004). It is also successfully applied to many high-dimensional procedures to resolve the dimensionality problem. For example, Schäfer and Strimmer (2005) reconstruct a gene regulatory network from microarray gene expression data using the inverse of a regularized covariance matrix. Chen et al. (2011) propose a modified Hotelling's T^2-statistic for testing high dimensional mean vectors and apply it to finding differentially expressed gene sets. We are motivated by the success of above examples and inspect whether the power can be improved by the reduced variability via linear shrinkage. To the best of our knowledge, our work is the first time to apply the linear shrinkage to the covariance matrix testing problem itself.

We derive the asymptotic distribution of the proposed $\mathrm{rLRT}(\lambda)$ under two scenarios, (i) when $\Sigma = \mathrm{I}_p$ for the null distribution, and additionally (ii) when $\Sigma = \Sigma_{\mathrm{spike}}$ for power study. Here Σ_{spike} means a covariance matrix from the spiked population model (Johnstone, 2001), roughly it is defined as a covariance matrix whose eigenvalues are all 1's but some finite nonunit 'spike'. The spiked covariance matrix assumed here includes the well known compound symmetry matrix $\Sigma_{\mathrm{cs}}(\rho) = \mathrm{I}_p + \rho \mathrm{J}_p$, where J_p is the $p \times p$ matrix of ones. The main results show that $\mathrm{rLRT}(\lambda)$ has normal distribution in asymptotic under both (i) and (ii); their asymptotic means are different but the variances are same. The main results are useful in testing various one sample covariance matrices. To be specific, first, in testing $\mathcal{H}_0 : \Sigma = \mathrm{I}_p$, (i) provides the asymptotic null distribution of $\mathrm{rLRT}(\lambda)$. Second, combining (i) and (ii) provides the asymptotic power for an arbitrary spiked alternative covariance matrix including $\Sigma_{\mathrm{cs}}(\rho)$. Finally, the results with $\lambda = 1$ provide various asymptotic distributions of the corrected LRT. Among these many applications, in this paper, we particularly focus on the LRT for testing $\mathcal{H}_0 : \Sigma = \mathrm{I}_p$, which has long been studied by many researchers (Anderson, 2003; Bai et al., 2009; Chen et al., 2010; Jiang et al., 2012; Ledoit and Wolf, 2002).

The paper is organized as follows. In Section 2, we briefly review results of the random matrix theory that are essential to the asymptotic theory of the proposed rLRT. The results include the limit of empirical spectral distribution (ESD) of the sample covariance matrix and the central limit theorem (CLT) for linear spectral statistics (LSS). In Section 3, we formally define the rLRT, and prove the asymptotic normality of the rLRT when the true

covariance matrix Σ is I_p or Σ_{spike}. In Section 4, the results developed in Section 3 are applied to testing $\mathcal{H}_0 : \Sigma = \mathrm{I}_p$. Numerical study is provided to compare the powers of the LRT and other existing methods including the corrected LRT and other non-LRT tests by Ledoit and Wolf (2002) and Chen et al. (2010). In Section 5, we conclude the paper with discussions of several technical details of the rLRT, for example, close spiked eigenvalues.

2 Random matrix theory

In this section, some useful properties of linear spectral statistics of the sample covariance matrix are introduced. The true covariance matrix Σ is identity or that from a spiked population model.

The following notation is used throughout the paper. Let M be a real-valued symmetric matrix of size $p \times p$ and $\alpha_j(\mathrm{M})$ be the $j-$th largest eigenvalue of the matrix M with natural labeling $\alpha_p(\mathrm{M}) \leq \cdots \leq \alpha_1(\mathrm{M})$. The *spectral distribution* (SD) for M is defined by

$$F^{\mathrm{M}}(t) := \frac{1}{p} \sum_{j=1}^{p} \delta_{\alpha_j(\mathrm{M})}(t), \quad t \in \mathbb{R},$$

where and $\delta_\alpha(t)$ is a point mass function that can be also written, with notational abuse, as $\delta_\alpha(t) = I(\alpha \leq t)$. Here, $I(A)$ denotes the indicator function of a set A.

2.1 Limiting spectral distribution of sample covariance matrix

Let $\{z_{ij}\}_{i,j\geq 1}$ be an infinite double array of independent and identically distributed real-valued random variables with $Ez_{11} = 0$, $Ez_{11}^2 = 1$ and $Ez_{11}^4 = 3$. Let $\mathbf{Z}_n = \{z_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, p\}$ be the top-left $n \times p$ block of the infinite double array. We assume that both n and p diverge and their ratio $\gamma' := p/n$ converges to a positive constant γ. The data matrix and the uncentered sample covariance matrix are $\mathbf{X}_n = \mathbf{Z}_n \Sigma_p^{1/2}$ and $\mathbf{S}_n^0 = \frac{1}{n} \mathbf{X}_n^\top \mathbf{X}_n$ respectively, where $\{\Sigma_p, p = 1, 2, \ldots\}$ is a sequence of $p \times p$ nonrandom symmetric matrices. Note that the fourth moment condition $Ez_{11}^4 = 3$ is used later on in Proposition 1.

In the random matrix theory literature (Bai et al., 2009; Bai and Silverstein, 2004), the limiting distribution of empirical SD $F^{\mathbf{S}_n^0}(\cdot)$ is determined by both the limits of p/n and $F^{\Sigma_p}(\cdot)$. Specifically, if $H_p(\cdot) := F^{\Sigma_p}(\cdot)$ converges in distribution to $H(\cdot)$, a random

distribution function $F^{\mathbf{S}_n^0}(\cdot)$ converges in distribution to a nonrandom distribution function, say $F^{\gamma,H}(\cdot)$ with probability one. The definition of $F^{\gamma,H}(\cdot)$ is given by its Stieltjes transform $m^{\gamma,H}(z)$ that is the unique solution of the following system of equations:

$$m^{\gamma,H}(z) = \frac{1}{\gamma}\underline{m}^{\gamma,H}(z) + \frac{1-\gamma}{\gamma z}; \tag{3}$$

$$z = -\frac{1}{\underline{m}^{\gamma,H}(z)} + \gamma \int \frac{t}{1+t\underline{m}^{\gamma,H}(z)} dH(t) \tag{4}$$

on $z \in \{z : \mathrm{Im}(z) > 0\}$. Generally, $\underline{m}^{\gamma,H}(z)$ is known as the Stieltjes transform of the limiting SD of the so-called companion matrix for $\mathbf{S}_n^0$, which is defined by $\underline{\mathbf{S}}_n^0 := \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^\top$. The density of $F^{\gamma,H}$ can be calculated from $\underline{m}^{\gamma,H}$ by the inversion formula,

$$\frac{dF^{\gamma,H}}{dx}(x) = \lim_{z\to x} \frac{1}{\gamma\pi} \, \mathrm{Im}[\underline{m}^{\gamma,H}(z)], \quad x \in \mathbb{R}, \quad z : \mathrm{Im}(z) > 0. \tag{5}$$

In the special case $\Sigma_p = \mathrm{I}_p$, we have $H_p(t) = \delta_1(t)$ and the corresponding spectral distribution F^{γ,δ_1} follows the Marčenko-Pastur law. To see this, note that the second equation of (4) can be rewritten as

$$z = -\frac{1}{\underline{m}^{\gamma,\delta_1}(z)} + \frac{\gamma}{1+\underline{m}^{\gamma,\delta_1}(z)}, \tag{6}$$

when $H = \delta_1$. By the inversion formula, (5) yields the probability density function of the Marčenko-Pastur distribution indexed by γ when $0 < \gamma < 1$,

$$\frac{dF^{\gamma,\delta_1}}{dx}(x) = \frac{1}{2\pi\gamma x}\sqrt{\big(b(\gamma)-x)(x-a(\gamma)\big)}, \quad a(\gamma) \le x \le b(\gamma), \tag{7}$$

where $a(\gamma) := (1-\sqrt{\gamma})^2$ and $b(\gamma) := (1+\sqrt{\gamma})^2$.

2.2 Central limit theorem for linear spectral statistics

Many multivariate statistical procedures are based on $F^{\mathbf{S}_n}$, the empirical SD of the centered sample covariance matrix $\mathbf{S}_n$. Consider is a family of functionals of eigenvalues that is also called as linear spectral statistics (LSS) or linear eigenvalue statistics:

$$\frac{1}{p}\sum_{j=1}^{p} g(l_j) = \int g(x)dF^{\mathbf{S}_n}(x)$$

where g is a function fulfilling certain complex-analytic conditions.

If the sample covariance matrix is uncentered, the central limit theorem (CLT) for the corresponding LSS is developed in Bai and Silverstein (2004) and Bai et al. (2009). The proposition below is adapted from Theorem 2.1 in Bai et al. (2009) and used in the asymptotic property of the proposed regularized LRT under the null. Note that the centering term of the CLT possesses a *finite-dimensional proxy* $F^{\gamma',H_p}(\cdot)$.

Proposition 1 (Bai et al. (2009))**.** *Let $T_n(g)$ be the functional*

$$T_n(g) = p\int g(x)\, d\Big\{F^{\mathbf{S}_n^0}(x) - F^{\gamma',H_p}(x)\Big\}\,. \tag{8}$$

Suppose that the two functions g_1 and g_2 , are complex analytic on an open domain containing an closed interval $[a(\gamma), b(\gamma)]$ on the real axis. If $\gamma' = p/n \to \gamma \in (0,1)$ and H_p converges in distribution to δ_1, then the vector $(T_n(g_1), T_n(g_2))$ converges in distribution to a bivariate normal distribution with mean

$$\mu(g_i) \;=\; \frac{g_i(a(\gamma)) + g_i(b(\gamma))}{4} - \frac{1}{2\pi}\int_{a(\gamma)}^{b(\gamma)} \frac{g_i(x)}{\sqrt{4\gamma - (x-1-\gamma)^2}}dx \tag{9}$$

for $i = 1, 2,$ and variance

$$v(g_1, g_2) = -\frac{1}{2\pi^2}\oint\oint \frac{g_1(z_1)g_2(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} d\underline{m}(z_1)d\underline{m}(z_2), \tag{10}$$

where $\underline{m} = \underline{m}^{\gamma,\delta_1}$ is defined in (6). The two contours in (10) are non-overlapping and containing $[a(\gamma), b(\gamma)]$, the support of F^{γ,δ_1}.

The proposition requires the data to have known population mean vector (known as zero without loss of generality) and considers the non-centered sample covariance matrix $\mathbf{S}_n^0$. However, this is seldomly true in practice and it is common to use the "unbiased" and "centered" sample covariance matrix $\mathbf{X}_n = \mathrm{B}_n\mathbf{Z}_n\Sigma_p^{1/2}$ and $\mathbf{S}_n = (1/\tilde{n})\mathbf{X}_n^\top\mathbf{X}_n$, where $\tilde{n} = n-1$ and $\mathrm{B}_n = \mathrm{I}_n - (1/n)\mathrm{J}_n$. The extension of Proposition 1 to the centered sample covariance matrix is studied recently in Zheng et al. (2015). At the paper, the authors prove that, under the fourth moment assumptions (the assumption $Ez_{11}^4 = 3$ in Section 2.1), Proposition 1 is still valid for $\mathbf{S}_n$ if $T_n(g)$ is redefined by

$$T_n(g) = p\int g(x)\, d\Big\{F^{\mathbf{S}_n}(x) - F^{\tilde{\gamma}',H_p}(x)\Big\}\,, \tag{11}$$

where $\tilde{n} = n - 1$ is the adjusted sample size and $\tilde{\gamma}' = p/\tilde{n}$. This is named as "substitution principle".

The assumption $H = \delta_1$ roughly implies that the spectrum of Σ_p is eventually concentrated around one. One simple example is $\Sigma_p = \mathrm{I}_p$. Then the SD is $H_p = F^{\mathrm{I}_p} = \delta_1$, which trivially converges to δ_1. In addition, we note that the spiked population model (Johnstone, 2001) has $H = \delta_1$ as the limiting SD and is applicable to Proposition 1. In our settings, the spiked population model refers to the data whose covariance matrix Σ_p has the following eigenvalue structure:

$$\underbrace{a_1, \ldots, a_1}_{n_1}, \underbrace{a_2, \ldots, a_2}_{n_2}, \ldots, \underbrace{a_k, \ldots, a_k}_{n_k}, \underbrace{1, \ldots, 1}_{p-K}. \tag{12}$$

Then the SD H_p corresponding to Σ_p is

$$H_p(t) = \frac{p-K}{p}\delta_1(t) + \frac{1}{p}\sum_{i=1}^{K} n_i \delta_{a_i}(t), \tag{13}$$

where $K = n_1 + \ldots + n_k$ is a fixed finite integer, not depending on n so that $p - K$ eigenvalues of unity eventually dominate corresponding H_p when p is large. Thus, the limiting SD remains unchanged as $H = \delta_1$.

To study the power of the proposed regularized LRT, it is useful to study how to apply the spiked population model to Proposition 1. Although the limiting SD is simple (δ_1), the spiked population model has several difficulties with the use of Proposition 1. In the spiked model, its SD, H_p in (13), has masses at $K+1$ distinct points and $\underline{m}^{\tilde{\gamma}', H_p}(z)$ is the solution to a polynomial equation of degree $K+2$. A polynomial equation has an analytical solution only when its degree is less than or equal to 4. Therefore, if $K \geq 3$, we do not have an analytic form of $\underline{m}^{\tilde{\gamma}', H_p}(z)$. To resolve this difficulty, recently, Wang et al. (2014) provides an approximation formula of $\int g dF^{\gamma', H_p}$ in Proposition 1 for the spiked population model. Such an approximation of $g(\underline{m}^{\gamma', H_p})$ at $\underline{m}^{\gamma', \delta_1}$ is constructed based on the idea that $\underline{m}^{\gamma', H_p}(z)$ and $\underline{m}^{\gamma', \delta_1}(z)$ would be close enough if p is large.

Proposition 2 (Wang et al. (2014)). *Suppose that $\gamma' < 1$ and H_p is given by (13), with $|a_i - 1| > \sqrt{\gamma}$ for all $i = 1, 2, \ldots, K$. If a complex-valued function g is analytic on an open domain containing the interval $[a(\gamma), b(\gamma)]$ and k points $\varphi(a_i) := a_i + \frac{\gamma a_i}{a_i - 1}$, $i = 1, 2, \ldots, K$*

on the real axis, then $\int g(x)dF^{\gamma',H_p}(x)$ *in (8) can be approximated by*

$$\begin{aligned}
&\int g(x)dF^{\gamma',H_p}(x)\\
&= -\frac{1}{2\pi ip}\oint_{\mathcal{C}} g(-\frac{1}{\underline{m}}+\frac{\gamma'}{1+\underline{m}})\left(\frac{K}{\gamma'\underline{m}}-\sum_{i=1}^{K}\frac{n_i a_i^2\underline{m}}{(1+a_i\underline{m})^2}\right)d\underline{m} && (14)\\
&+\frac{1}{2\pi ip}\oint_{\mathcal{C}} f'(-\frac{1}{\underline{m}}+\frac{\gamma'}{1+\underline{m}})\sum_{i=1}^{K}\frac{(1-a_i)n_i}{(1+a_i\underline{m})(1+\underline{m})}\left(\frac{1}{\underline{m}}-\frac{\gamma'\underline{m}}{(1+\underline{m})^2}\right)d\underline{m} && (15)\\
&+\left(1-\frac{K}{p}\right)\int g(x)dF^{\gamma',\delta_1}(x)+\frac{1}{p}\sum_{i=1}^{K}n_i g(\varphi(a_i))+O(\frac{1}{n^2}) && (16)
\end{aligned}$$

where $\underline{m}=\underline{m}^{\gamma',\delta_1}$ *is defined in (6) by substituting* γ *by* γ'*, and* $\mathcal{C}$ *is a counterclockwise contour enclosing the interval* $\left[\frac{-1}{1-\sqrt{\gamma'}},\frac{-1}{1+\sqrt{\gamma'}}\right]$ *on the real axis.*

The above proposition is a special case of Theorem 2 in Wang et al. (2014) when all the a_i's are *distant spikes*, i.e., $|a_i-1|>\sqrt{\gamma}$. Note that Theorem 2 of Wang et al. (2014) allows *close spikes* a_i that is defined by $|a_i-1|\le\sqrt{\gamma}$. In this paper, we focus on the alternative hypothesis with distant spike in the power study. We finally remark that the substitution principle is directly applicable to Wang et al.'s results, that is, one can approximate $\int g(x)dF^{\tilde{\gamma}',H_p}(x)$ in (11) by Proposition 2, with $\gamma'=p/n$ in the formula replaced by $\tilde{\gamma}'=p/(n-1)$.

3 Main results

In this section, the asymptotic results of the rLRT are presented. Here, the rLRT is defined via the linear shrinkage estimator instead of the sample covariance matrix :

$$\mathrm{rLRT}(\lambda):=\mathrm{tr}\big(\widehat{\Sigma}\big)-\log\big|\widehat{\Sigma}\big|-p,\quad\text{where }\widehat{\Sigma}:=\lambda\mathbf{S}_n+(1-\lambda)\mathrm{I}_p.$$

The shrinkage intensity λ is fixed and chosen from $(0,1)$. Define $\psi(x)=\lambda x+(1-\lambda)$ and $g(x)=\psi(x)-\log\{\psi(x)\}-1$. We consider $\mathrm{rLRT}(\lambda)=p\int g(x)dF^{\mathbf{S}_n}(x)$, whose sample covariance matrix term is of the centered version. Then Proposition 1 along with the substitution principle yields the following results.

Theorem 1. *Let* $g(x)=\psi(x)-\log\{\psi(x)\}-1$ *and* $\psi(x)=\lambda x+(1-\lambda)$ *with fixed* $\lambda\in(0,1)$*. Suppose that* $\Sigma_p=\mathrm{I}_p$*. If* $\tilde{\gamma}'=p/\tilde{n}\to\gamma\in(0,1)$ *with* $\tilde{n}=n-1$*, then*

$$T_n(g)=\mathrm{rLRT}(\lambda)-p\int g(x)dF^{\tilde{\gamma}',\delta_1}(x)$$

converges in distribution to the normal distribution with mean

$$\mu(g) = -\frac{\log\sqrt{(1+\lambda\gamma)^2 - 4\lambda^2\gamma}}{2} + \frac{1}{4\pi}\int_0^{2\pi} \left(\log(1+\lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)\right) d\theta \tag{17}$$

and variance

$$v(g) = 2\left\{-\frac{\lambda}{M} - \lambda(1+\gamma-\lambda\gamma) + \frac{\lambda\gamma}{1+N} - \log\frac{M-N}{M(1+N)}\right\}, \tag{18}$$

where

$$M, N = M(\lambda,\gamma), N(\lambda,\gamma) := \frac{-(1-2\lambda+\lambda\gamma) \pm \sqrt{(1-2\lambda+\lambda\gamma)^2 + 4\lambda(1-\lambda)}}{2(1-\lambda)} \tag{19}$$

The detailed proof of Theorem 1 is given in Appendix A.

Note that when $\lambda = 1$, $\mu(g) = -\log(1-\gamma)/2$ and $v(g) = -2\gamma - 2\log(1-\gamma)$ that are consistent with the results in Bai et al. (2009). To see this, observe that the integral in the mean function $\int_0^{2\pi} \log\left(1+\lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta\right) d\theta$ approaches zero according to the dominate convergence theorem. In addition, it can be shown that M goes to $-1/(1-\gamma)$, and N goes to $+\infty$ as $\lambda \to 1$ using the approximation formula $\sqrt{x+\Delta x} \approx \sqrt{x} + \frac{1}{2}x^{-1/2}\Delta x$.

Next, consider the finite-dimensional proxy

$$\int g(x) dF^{\tilde{\gamma}', \delta_1}(x).$$

From the density function of Marčenko-Pastur law (7) and the fact that $1 = \int x dF^{\tilde{\gamma}', \delta_1}(x) = \int 1 dF^{\tilde{\gamma}', \delta_1}(x)$,

$$\begin{aligned}
& \int g(x) dF^{\tilde{\gamma}', \delta_1}(x) \\
= & \int_{a(\tilde{\gamma}')}^{b(\tilde{\gamma}')} \frac{(\lambda x - \lambda) - \log(\lambda x + 1 - \lambda)}{2\pi x \tilde{\gamma}'} \sqrt{\{b(\tilde{\gamma}') - x\}\{x - a(\tilde{\gamma}')\}} dx \\
= & -\int_{a(\tilde{\gamma}')}^{b(\tilde{\gamma}')} \frac{\log(\lambda x + 1 - \lambda)}{2\pi x \tilde{\gamma}'} \sqrt{\{b(\tilde{\gamma}') - x\}\{x - a(\tilde{\gamma}')\}} dx.
\end{aligned}$$

By substituting $x = 1 + \tilde{\gamma}' - 2\sqrt{\tilde{\gamma}'}\cos\theta$, we have an alternative representation of the integral

$$\int g(x) dF^{\tilde{\gamma}', \delta_1}(x) = -\frac{2}{\pi}\int_0^{\pi} \frac{\log(1+\lambda\tilde{\gamma}' - 2\lambda\sqrt{\tilde{\gamma}'}\cos\theta)}{1+\tilde{\gamma}' - 2\sqrt{\tilde{\gamma}'}\cos\theta} \cdot \sin^2\theta d\theta. \tag{20}$$

It is remarked that the right-hand-side of (20) can be evaluated via the standard numerical integration techniques.

Theorem 2 below establishes the asymptotic normality of the rLRT under the alternative hypothesis that the true covariance matrix from the spiked population model in Section 2. It follows directly from Proposition 2.

Theorem 2. *Let $g(x) = \psi(x) - \log\{\psi(x)\} - 1$ and $\psi(x) = \lambda x + (1-\lambda)$ with fixed $\lambda \in (0,1)$. Suppose that Σ_p has SD $H_p(t) = \frac{p-K}{p}\delta_1(t) + \frac{1}{p}\sum_{i=1}^K n_i \delta_{a_i}(t)$ as in (13) with $|a_i - 1| > \sqrt{\gamma}$ for all $i = 1, 2, \ldots, K$. If $\tilde{\gamma}' = p/\tilde{n} \to \gamma \in (0,1)$ with $\tilde{n} = n-1$, then*

$$\mathrm{rLRT}(\lambda) - p\int g(x) dF^{\tilde{\gamma}', H_p}(x) \longrightarrow N\big(\mu(g), v(g)\big),$$

where $\mu(g), v(g)$ are defined in Theorem 1 and

$$p\int g(x) dF^{\tilde{\gamma}', H_p}(x) = \left(1 - \frac{K}{p}\right)\int g(x) dF^{\tilde{\gamma}', \delta_1}(x) + \frac{K}{p} C(\lambda, \tilde{\gamma}') + O\left(\frac{1}{n^2}\right)$$

with

$$\begin{aligned} C(\lambda, \tilde{\gamma}') \;=\; & \lambda \cdot \frac{1}{K}\sum_{i=1}^K n_i a_i - \lambda - \frac{1}{K}\sum_{i=1}^K n_i \log \psi\{\varphi(a_i)\} \\ & - \left[\frac{1}{\tilde{\gamma}'}\log(-M) + \frac{1}{K}\sum_{i=1}^K n_i \log\left(\frac{1-a_i}{1+a_i M}\right) - \frac{1}{K}\sum_{i=1}^K n_i \left(\frac{1}{1+a_i M} - \frac{1}{1-a_i}\right)\right] \\ & + \frac{\lambda}{(1-\lambda)K}\sum_{i=1}^K n_i \left[\frac{1}{M-N}\left\{\frac{a_i(M+1)}{1+a_i M} - \frac{a_i \tilde{\gamma}' M^2}{(1+a_i M)(M+1)} - 1 + \frac{\tilde{\gamma}' M^2}{(M+1)^2}\right\}\right. \\ & \left. - \frac{1}{(M+1)(N+1)}\left\{\frac{a_i \tilde{\gamma}'}{1-a_i} + \frac{\tilde{\gamma}'(2MN + M + N)}{(M+1)(N+1)}\right\}\right], \end{aligned}$$

where $\varphi(a) = a + \frac{\gamma a}{a-1}$ and both M and N in $C(\lambda, \tilde{\gamma}')$ are $M = M(\lambda, \tilde{\gamma}')$ and $N = N(\lambda, \tilde{\gamma}')$.

The proof of the above theorem is provided in Appendix A.

Theorem 2 can be applied to Σ_{cs}, the covariance matrix with compound symmetry, which is defined by

$$\Sigma_{\mathrm{cs}}\left(\frac{\beta}{p}\right) = \begin{pmatrix} 1+\beta/p & \beta/p & \beta/p & \cdots & \beta/p \\ \beta/p & 1+\beta/p & & & \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \beta/p & \cdots & \cdots & \cdots & 1+\beta/p \end{pmatrix} = \mathrm{I}_p + \frac{\beta}{p}\mathrm{J}_p. \tag{21}$$

This matrix has a spiked eigenvalue structure; $1+\beta$ for one eigenvalue and 1 for the other $p-1$ eigenvalues. The corresponding SD is $H_p(t) = \frac{p-1}{p}\delta_1(t) + \frac{1}{p}\delta_{1+\beta}(t)$. Theorem 2 with $K=1$, $k=1$, $n_1 = 1$, and $a_1 = 1+\beta$ gives the following corollary.

Corollary 1. *Let $g(x) = \psi(x) - \log\{\psi(x)\} - 1$ and $\psi(x) = \lambda x + (1-\lambda)$ with fixed $\lambda \in (0,1)$. Suppose that Σ_p has SD $H_p(t) = \frac{p-1}{p}\delta_1(t) + \frac{1}{p}\delta_{1+\beta}(t)$ with $\beta > \sqrt{\gamma}$. If $\tilde{\gamma}' := p/\tilde{n} \to \gamma \in (0,1)$ where $\tilde{n} = n-1$, then*

$$\mathrm{rLRT}(\lambda) - p\int g(x)dF^{\tilde{\gamma}',H_p}(x) \longrightarrow N(\mu(g), v(g)),$$

where $\mu(g), v(g)$ are defined in Theorem 1 and

$$\int g(x)dF^{\tilde{\gamma}',H_p}(x) = \left(1 - \frac{1}{p}\right)\int g(x)dF^{\tilde{\gamma}',\delta_1}(x) + \frac{1}{p}C(\lambda,\tilde{\gamma}') + O\left(\frac{1}{n^2}\right)$$

with

$$\begin{aligned} C(\lambda,\tilde{\gamma}') &= \lambda\beta - \log\psi\{\varphi(1+\beta)\} + \frac{1}{\tilde{\gamma}'}\left\{\frac{1}{\lambda} + \log(-\frac{1}{M})\right\} \\ &\quad + \frac{1}{1+(1+\beta)M} - \log\left(-\frac{\beta}{1+(1+\beta)M}\right) \\ &\quad + \frac{\lambda}{(1-\lambda)(M-N)}\left\{\frac{(1+\beta)(M+1)}{1+(1+\beta)M} - \frac{\tilde{\gamma}'(1+\beta)M^2}{(1+(1+\beta)M)(M+1)} - 1 + \frac{\tilde{\gamma}'M^2}{(M+1)^2}\right\}, \end{aligned}$$

where $\varphi(a) = a + \frac{\gamma a}{a-1}$ and both M and N in $C(\lambda,\tilde{\gamma}')$ are $M = M(\lambda,\tilde{\gamma}')$ and $N = N(\lambda,\tilde{\gamma}')$.

In the corollary, the condition $\beta > \sqrt{\gamma}$ means that the spiked eigenvalue $1+\beta$ is distant.

4 Testing the identity covariance matrix

The results in Section 3 can be used for testing various hypotheses on one sample covariance matrix. In this section, we study the finite-sample properties of the proposed rLRT in testing $\mathcal{H}_0 : \Sigma = \mathrm{I}_p$. Additionally, we compare the power of the proposed rLRT and the following existing procedures in the literature.

- Ledoit and Wolf (2002) assume that $p/n \to \gamma \in (0,\infty)$ and propose a statistic

$$\mathbf{T}_{\mathrm{LW}} = \frac{1}{p}\mathrm{tr}\left\{(\mathbf{S} - \mathrm{I})^2\right\} - \frac{p}{n}\left\{\frac{1}{p}\mathrm{tr}(\mathbf{S})\right\}^2 + \frac{p}{n}.$$

  The asymptotic distribution of $\mathbf{T}_{\mathrm{LW}}$ is, if both n and p increase with $p/n \to \gamma \in (0,\infty)$,

$$n\mathbf{T}_{\mathrm{LW}} - p$$

  converges in distribution to normal distribution with mean 1 and variance 4.

- Bai et al. (2009) and Zheng et al. (2015) propose a corrected LRT for the cases where both n and p increases and p/n converges to $\gamma \in (0,1)$. The corrected LRT statistic is

$$\text{cLRT} = \operatorname{tr}\big(\mathbf{S}_n\big) - \log\big|\mathbf{S}_n\big| - p.$$

They show that

$$\mathbf{T}_{\text{lrt}} = v(g)^{-1/2}\left\{\text{cLRT} - p\int g(x)dF^{\tilde{\gamma}',\delta_1}(x)dx - \mu(g)\right\}$$

converges in distribution to the standard normal distribution, where $\mu(g) = -\frac{\log(1-\gamma)}{2}$, $v(g) = -2\log(1-\gamma) - 2\gamma$, and $\int g(x)dF^{\tilde{\gamma}',\delta_1}(x)dx = 1 - \frac{\tilde{\gamma}'-1}{\tilde{\gamma}'}\log\big(1-\tilde{\gamma}'\big)$.

- Finally, Chen et al. (2010) proposes to use the statistic

$$\mathbf{T}_{\text{C}} = \frac{1}{p}\mathbf{V}_{2.n} - \frac{2}{p}\mathbf{V}_{1.n} + 1,$$

where

$$\begin{aligned}
\mathbf{V}_{1.n} &= \frac{1}{n}\sum_{i=1}^{n} X_i^\top X_i - \frac{1}{P_n^2}\sum_{i\neq j} X_i^\top X_j \\
\mathbf{V}_{2.n} &= \frac{1}{P_n^2}\sum_{i\neq j}\big(X_i^\top X_j\big)^2 - 2\frac{1}{P_n^3}\sum_{i,j,k}^{*} X_i^\top X_j X_j^\top X_k \\
&\quad + \frac{1}{P_n^4}\sum_{i,j,k,l}^{*} X_i^\top X_j X_k^\top X_l \\
P_n^r &= n!/(n-r)!
\end{aligned}$$

and $\sum^*$ is the summation over different indices. The asymptotic theory suggests that, under the null, $n\mathbf{T}_{\text{C}}$ converges in distribution to the normal distribution with mean 0 and variance 4.

4.1 Power comparison with the cLRT

The asymptotic power curves of the cLRT and rLRT for the alternative hypothesis of the compound symmetry can be obtained using Corollary 1. When $\Sigma_p \equiv \Sigma_{\text{cs}}(\beta/p)$ and $\tilde{\gamma}' \to \gamma$, the probability of rejecting $\mathcal{H}_0 : \Sigma_p = \text{I}_p$ at level η is

$$1 - \Phi\left[\Phi^{-1}(\eta) + \frac{1}{v(g)}\left(\int g(x)dF^{\tilde{\gamma}',\delta_1}(x) - C(\lambda,\tilde{\gamma}')\right)\right], \quad \beta > \sqrt{\gamma} \tag{22}$$

where $C(\lambda, \tilde{\gamma}')$ is defined in Corollary 1 and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

The powers of the cLRT and rLRT with $\lambda = 0.4$ and $\lambda = 0.7$ are plotted in Figure 1. Each panel of Figure 1 compares the powers of the cLRT and rLRT for different sample size n. In each panel, the results of $\beta < \sqrt{\gamma}$ (close spike) are also included to study the performances when the the assumption of distant spike in Corollary 1 is violated. The results in Figure 1 suggest that Theorem 2 would be applicable when there is a "close spike" eigenvalue. More detailed discussions are given in Section 5.

We find that in all cases the rLRT has higher empirical power than the cLRT for the chosen values of $\lambda = 0.4$ and 0.7. We also find that the empirical power curve increases to 1 less rapidly if λ or γ is closer to 1. In addition, although we do not report the details, the empirical curves converge fast and do have minor changes after $n = 80$ for the selected values of λ and γ.

To understand the power gain due to the use of the rLRT better, we plot the empirical density of the rLRT and cLRT under the null and four alternative hypotheses (A1)-(A4) (used in Section 4.2) in Figure 2. The figure shows that (i) the variances of the rLRT are smaller than the cLRT under both null and alternative hypotheses and (ii) the distances between the null and the alternative distributions are larger in the rLRT than in the cLRT. Accordingly, the rLRT has larger power than the cLRT and we will see this is true regardless of the choice of n and p in the next section.

4.2 Power comparison with other existing procedures

In this section, we numerically compare the empirical sizes and powers of the rLRT statistic to other existing tests, namely the corrected LRT (cLRT) by Bai et al. (2009), the invariant test by Ledoit and Wolf (2002), and the non-parametric test by Chen et al. (2010).

In this study, random samples of size n are generated from the $p-$dimensional multivariate normal distribution $\mathrm{MVN}_p(0, \Sigma)$. The covariance matrix is set as $\Sigma = \mathrm{I}_p$ to obtain the empirical sizes. The sample size n is chosen as $20, 40, 80$, and 160, and, for each n, $\gamma = \gamma' = p/n$ is chosen as $0.2, 0.5$, and 0.8. For example, $p = 32, 80$, and 128 are considered for $n = 160$ in the simulation. We take 0.05 as as the level of significance. The shrinkage intensity λ of $\mathrm{rLRT}(\lambda)$ is selected from $0.2, 0.5$, and 0.8 to investigate the effect of the magnitude of

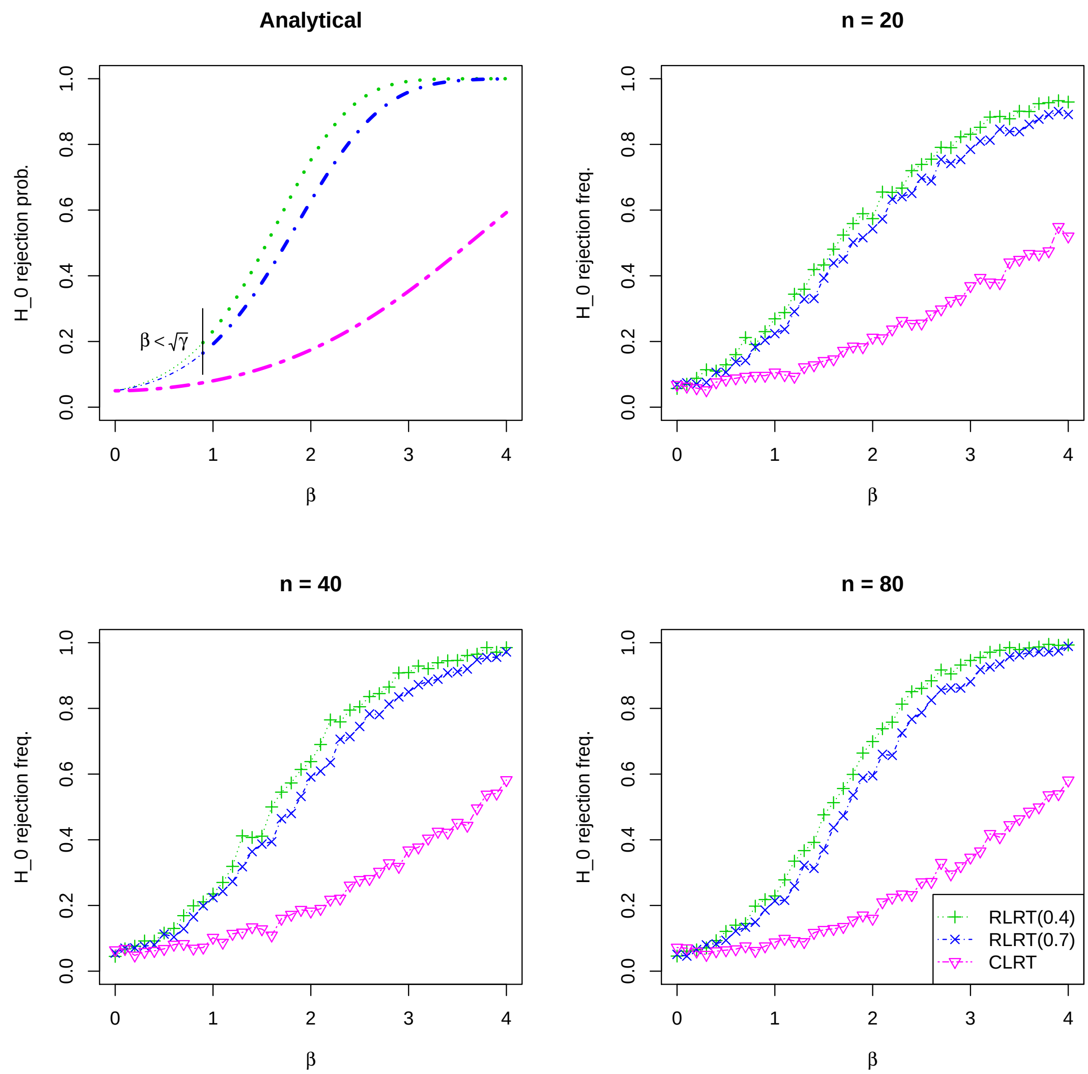

Figure 1: Analytic and empirical power curves for the rLRT and cLRT.

the linear shrinkage. This means that we compare the empirical sizes of cLRT, rLRT(0.2), rLRT(0.5), and rLRT(0.8) under varying n and γ.

To compare the powers, we consider the following four alternatives: (A1) independent but heteroscedastic variance case $\Sigma = \mathrm{diag}(2, 2, \cdots, 2, 1, 1, \cdots, 1)$ where the number of 2's is $\min\{1, \lfloor 0.2p \rfloor\}$, where $\lfloor r \rfloor$ is the round-down of r; (A2) independent with a single diverging spiked eigenvalue as $\Sigma = \mathrm{diag}(1 + 0.2p, 1, 1, \cdots, 1)$; (A3) compound symmetry $\Sigma = \Sigma_{\mathrm{cs}}(0.2)$; and (A4) compound symmetry $\Sigma = \Sigma_{\mathrm{cs}}(0.1)$ as defined in (21). Here, the compound symmetry matrix $\Sigma_{\mathrm{cs}}(\rho)$ has a single spiked eigenvalue $1 + \rho \cdot p$ and $p - 1$ non-spiked eigenvalues

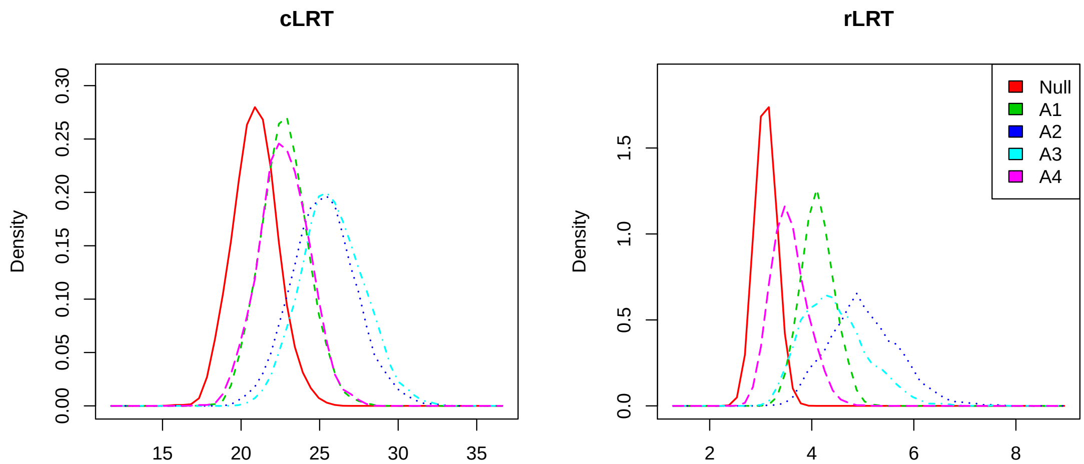

Figure 2: Empirical density functions of the rLRT and the cLRT under the null and four alternative hypotheses when $n = 40$, $p = 32$, and $\lambda = 0.5$: (A1) independent but heteroscedastic variance case $\Sigma = \text{diag}(2, 2, \cdots, 2, 1, 1, \cdots, 1)$ where the number of 2's is $\min\{1, \lfloor 0.2p \rfloor\}$, where $\lfloor r \rfloor$ is the round-down of r; (A2) independent with a single diverging spiked eigenvalue as $\Sigma = \text{diag}(1 + 0.2p, 1, 1, \cdots, 1)$; (A3) compound symmetry $\Sigma = \Sigma_{\text{cs}}(0.2)$; and (A4) compound symmetry $\Sigma = \Sigma_{\text{cs}}(0.1)$ as defined in (21).

of 1. Thus, (A2) and (A3) have the identical spectra. The sample size n and γ are chosen to be the same as those of the null.

The empirical sizes and powers of the listed methods are reported in Table 1. First, the empirical sizes of all the tests approach to the aimed level 0.05 as n increases. However, the size of Ledoit and Wolf (2002) shows slower convergence and more upward bias than the size of the other tests do in all cases we considered here. For this reason, the power of Ledoit and Wolf (2002) after correcting the size (empirically) are also reported in Table 1, where the cut off value is decided based on $100,000$ simulated test statistics under the null for each simulation setting. Second, it can be seen that the emprical powers of the rLRTs are higher than those of the cLRT in all cases we considered. In addition, it is interesting to note that the power improvement is especially higher in the case $\gamma = 0.8$ (when p is relatively largfe). Third, comparing to the tests of Chen et al. (2010) and (the biased-corrected version of) Ledoit and Wolf (2002), the proposed rLRT(0.5) and rLRT(0.2) has higher empirical power in most of the cases. Finally, we remark that the computational cost of Chen et al. (2010) is at least $O(pn^4)$ due to the fourth moment calculation so it is not suitable for data with

large n. In fact, to test the data with $n = 500$, Chen et al. (2010) takes tens of hours to finish all computation (using `C` codes and Intel Core-i7 CPU), whereas the other tests only require seconds.

5 Discussion

We conclude the paper with a few additional issues of the proposed rLRT not fully discussed in the mainbody of the paper.

First, we consider the case where $n < p$; equivalently, $\gamma \geq 1$. In this case, the cLRT is not well-defined because the logarithm term, $\log|\mathbf{S}_n| = \sum_i \log(l_i)$, contains some zero l_i's. On the other hand, the rLRT is still well-defined as the corresponding logartithm term $\sum_i \log\psi(l_i)$ remains positive even if $\lambda < 1$. Since the CLT of the linear spectral statistics holds for $\gamma \in [0, \infty)$ (Bai and Silverstein, 2004), it is possible to extend Theorem 1 and 2 in this paper to the case where $\gamma \in [1, \infty)$.

Second, we discuss the case with closely spiked eigenvalues. In the compound symmetry model of (21), the close spiked eigenvalues are those where the spike $1 + \beta$ is smaller than $1 + \sqrt{\gamma}$. As shown in Figure 1, it appears that the power curves over the interval $\beta \in (0, \sqrt{\gamma})$ could be obtained simply by extending the formula of Corollary 1 to $(0, \sqrt{\gamma})$. We remark that, however, if we follow Wang et al. (2014), the term $\log\psi(1 + \beta)$ in Corollary 1 should be omitted when $\beta \in (0, \sqrt{\gamma})$, leading to incoherence between the analytical and empirical power curves on $(0, \sqrt{\gamma})$.

Finally, the selection of shrinkage intensity λ is still not well understood for hypothesis testing. As a reviewer points out, when λ approaches 0, the rLRT becomes irrelevant with the alternative covariance matrix and its power is expected to be close to the size. Thus, an appropriate selection of λ is important for good performance of the rLRT. The selection of λ for the purpose of improved estimation is well studied in the literature, for example, Ledoit and Wolf (2004), Schäfer and Strimmer (2005) and Warton (2008). However, our additional numerical study shows that such a choice of λ for the estimation purpose cannot achieve a power gain in testing problem. The optimal selection of λ for the hypothesis testing needs further research.

	γ	n	cLRT	rLRT (0.8)	rLRT (0.5)	rLRT (0.2)	Chen	LW
Null	0.2	20	0.085	0.078	0.078	0.083	0.087	0.089
		40	0.069	0.066	0.068	0.073	0.074	0.083
		80	0.062	0.059	0.061	0.064	0.070	0.076
	0.5	20	0.068	0.059	0.065	0.072	0.068	0.104
		40	0.061	0.055	0.059	0.064	0.075	0.098
		80	0.056	0.053	0.056	0.060	0.068	0.091
	0.8	20	0.065	0.055	0.063	0.068	0.068	0.125
		40	0.058	0.052	0.056	0.059	0.058	0.120
		80	0.055	0.051	0.054	0.056	0.064	0.113
A1: Indep. with hetero. variance	0.2	20	0.370	0.444	0.511	**0.550**	0.459	0.517 (0.440)
		40	0.394	0.489	0.579	**0.632**	0.540	0.596 (0.527)
		80	0.951	0.984	0.996	**0.998**	0.989	0.991 (0.987)
	0.5	20	0.265	0.478	0.604	**0.653**	0.478	0.574 (0.471)
		40	0.569	0.880	0.961	**0.977**	0.851	0.897 (0.849)
		80	0.967	1.000	1.000	1.000	1.000	1.000 (0.999)
	0.8	20	0.180	0.580	0.703	**0.741**	0.459	0.629 (0.492)
		40	0.392	0.959	0.991	**0.995**	0.854	0.924 (0.868)
		80	0.859	1.000	1.000	1.000	0.999	1.000 (0.999)
A2: Indep. with single diversing spike $1 + 0.2p$	0.2	20	0.280	0.341	0.403	**0.440**	0.351	0.408 (0.331)
		40	0.724	0.810	0.871	**0.899**	0.853	0.882 (0.848)
		80	0.999	1.000	1.000	1.000	1.000	1.000 (1.000)
	0.5	20	0.385	0.596	0.698	**0.739**	0.664	0.729 (0.650)
		40	0.890	0.979	0.993	**0.996**	0.989	0.996 (0.994)
		80	1.000	1.000	1.000	1.000	1.000	1.000 (1.000)
	0.8	20	0.349	0.765	0.838	**0.864**	0.795	0.879 (0.816)
		40	0.867	0.997	0.999	**1.000**	0.999	1.000 (0.999)
		80	1.000	1.000	1.000	1.000	1.000	1.000 (1.000)
A3: Compound symmetry with $\rho = 0.2$	0.2	20	0.280	0.343	0.404	**0.440**	0.382	0.410 (0.333)
		40	0.726	0.810	0.870	**0.900**	0.835	0.882 (0.848)
		80	0.999	1.000	1.000	1.000	1.000	1.000 (1.000)
	0.5	20	0.387	0.595	0.695	**0.736**	0.669	0.727 (0.649)
		40	0.888	0.979	0.992	**0.996**	0.995	0.996 (0.993)
		80	1.000	1.000	1.000	1.000	1.000	1.000 (1.000)
	0.8	20	0.351	0.764	0.837	**0.862**	0.796	0.877 (0.816)
		40	0.865	0.997	0.999	1.000	0.999	1.000 (0.999)
		80	1.000	1.000	1.000	1.000	1.000	1.000 (1.000)
A4: Compound symmetry with $\rho = 0.1$	0.2	20	0.137	0.162	0.195	**0.217**	0.187	0.201 (0.140)
		40	0.280	0.358	0.441	**0.491**	0.400	0.454 (0.380)
		80	0.796	0.885	0.939	**0.962**	0.948	0.953 (0.940)
	0.5	20	0.153	0.252	0.330	**0.368**	0.292	0.365 (0.266)
		40	0.405	0.659	0.782	**0.830**	0.772	0.828 (0.767)
		80	0.939	0.996	0.999	1.000	1.000	1.000 (1.000)
	0.8	20	0.139	0.363	0.452	**0.490**	0.403	0.526 (0.392)
		40	0.376	0.838	0.911	**0.936**	0.912	0.952 (0.920)
		80	0.915	1.000	1.000	1.000	1.000	1.000 (1.000)

Table 1: Summary of sizes and powers over 100,000 replications except for Chen et al. (2010) and 1,000 replications for Chen et al. (2010) (due to heavy computation). The empirically corrected powers of Ledoit and Wolf (2002) are reported in the parentheses. For each row, the maximum powers are highlighted in bold and the maximum powers among the four LRT-based tests are underlined. The powers for $n = 160$ are all equal to 1 and are removed from the table.

Acknowledgements

This paper is supported by National Research Foundation of Korea (NRF) grant funded by the government (MSIP) (No. 2011-0030810). The authors are grateful to two reviewers and the AE for a careful reading and providing improvements.

A Proof of the main theorems

A.1 Proof of Theorem 1

Theorem 1 is a direct consequence of Proposition 1. Here, we calculate the integrals (9) and (10) for

$$g(z) = g_1(z) = g_2(z) = \psi(z) - \log(\psi(z)) - 1,$$

where $\psi(z) = \lambda z + (1 - \lambda)$. Mean: Using Proposition 1 and the substitution

$$x = 1 + \gamma - 2\sqrt{\gamma}\cos\theta\,,$$

$$\begin{aligned}
\mu(g_i) &= \frac{g_i(a(\gamma)) + g_i(b(\gamma))}{4} - \frac{1}{2\pi}\int_{a(\gamma)}^{b(\gamma)} \frac{g_i(x)}{\sqrt{4\gamma - (x-1-\gamma)^2}}dx \\
&= \frac{\lambda\gamma - \log\sqrt{(1+\lambda\gamma)^2 - 4\lambda^2\gamma}}{2} \\
&\quad -\frac{1}{2\pi}\int_0^{\pi} \left(\lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta - \log(1 + \lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)\right) d\theta \\
&= -\frac{\log\sqrt{(1+\lambda\gamma)^2 - 4\lambda^2\gamma}}{2} + \frac{1}{2\pi}\int_0^{\pi} \left(\log(1 + \lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)\right) d\theta \\
&= -\frac{\log\sqrt{(1+\lambda\gamma)^2 - 4\lambda^2\gamma}}{2} + \frac{1}{4\pi}\int_0^{2\pi} \left(\log(1 + \lambda\gamma - 2\lambda\sqrt{\gamma}\cos\theta)\right) d\theta\,.
\end{aligned}$$

Variance: We write $m_1 := \underline{m}(z_1)$ and $m_2 = \underline{m}(z_2)$ for notational simplicity. We have

$$v(g) = -\frac{1}{2\pi^2}\oint g(z_2(m_2))\oint \frac{g(z_1(m_1))}{(m_1 - m_2)^2}dm_1 dm_2, \tag{23}$$

To evaluate this integral with Cauchy's formula, we need to identify the points of singularity in $g(z(m_1))$. It can be seen that there is singularity when $\psi(z(m_1)) = 0$. Rewrite $\psi(z(m))$ as

$$\psi(z(m_1)) = \lambda\left(-\frac{1}{m_1} + \frac{\gamma}{m_1+1}\right) + 1 - \lambda = \frac{(1-\lambda)(m_1 - M)(m_1 - N)}{m_1(m_1+1)}$$

where

$$M, N = \frac{-(1-2\lambda+\lambda\gamma) \pm \sqrt{(1-2\lambda+\lambda\gamma)^2 + 4\lambda(1-\lambda)}}{2(1-\lambda)}.$$

Then, M, N are points of singularity. Next, choose contours $\mathcal{C}_1$ and $\mathcal{C}_2$ enclosing -1 and M, but not 0 and N, such that on the contours, the logarithm in $g(z)$ is single-valued. In addition, $\mathcal{C}_1$ and $\mathcal{C}_2$ are chosen so that they do not overlap. Applying integration by parts and Cauchy's formula, we have

$$\begin{aligned}
&\oint \frac{g(z(m_1))}{(m_1-m_2)^2} dm_1 \\
= &\oint \left\{ \frac{\lambda}{m_1^2} - \frac{\lambda\gamma}{(m_1+1)^2} + \frac{1}{m_1} + \frac{1}{m_1+1} - \frac{1}{m_1-M} - \frac{1}{m_1-N} \right\} \frac{1}{m_1-m_2} dm_1 \\
= &\ 2\pi i \left\{ \frac{\lambda\gamma}{(m_2+1)^2} - \frac{1}{m_2+1} + \frac{1}{m_2-M} \right\}
\end{aligned} \tag{24}$$

Then,

$$\begin{aligned}
v(g) &= -\frac{1}{2\pi^2} \oint g(z_2(m_2)) \oint \frac{g(z_1(m_1))}{(m_1-m_2)^2} dm_1 dm_2 \\
&= -\frac{i}{\pi} \oint g(z_2(m_2)) \left\{ \frac{\lambda\gamma}{(m_2+1)^2} - \frac{1}{m_2+1} + \frac{1}{m_2-M} \right\} dm_2 .
\end{aligned}$$

Here,

$$\begin{aligned}
&\oint \frac{g(z(m_2))}{(m_2+1)^2} dm_2 \\
= &\oint \left\{ \frac{\lambda}{m_2^2} - \frac{\lambda\gamma}{(m_2+1)^2} + \frac{1}{m_2} + \frac{1}{m_2+1} - \frac{1}{m_2-M} - \frac{1}{m_2-N} \right\} \frac{1}{m_2+1} dm_2 \\
= &\ \frac{2\pi i}{1+N}
\end{aligned} \tag{25}$$

Applying integration by parts, Cauchy's formula, and Lemma 1, we have

$$\begin{aligned}
&\oint \frac{g(z(m_1))}{(m_1-m_2)^2} dm_1 \\
= &\oint \left\{ \frac{\lambda}{m_1^2} - \frac{\lambda\gamma}{(m_1+1)^2} + \frac{1}{m_1} + \frac{1}{m_1+1} - \frac{1}{m_1-M} - \frac{1}{m_1-N} \right\} \frac{1}{m_1-m_2} dm_1 \\
= &\ 2\pi i \left\{ \frac{\lambda\gamma}{(m_2+1)^2} - \frac{1}{m_2+1} + \frac{1}{m_2-M} \right\}
\end{aligned} \tag{26}$$

Then,

$$\begin{aligned}
v(g) &= -\frac{1}{2\pi^2} \oint g(z_2(m_2)) \oint \frac{g(z_1(m_1))}{(m_1-m_2)^2} dm_1 dm_2 \\
&= -\frac{i}{\pi} \oint g(z_2(m_2)) \left\{ \frac{\lambda\gamma}{(m_2+1)^2} - \frac{1}{m_2+1} + \frac{1}{m_2-M} \right\} dm_2 .
\end{aligned}$$

Using integration by parts,

$$\begin{aligned}
&\oint \frac{g(z(m_2))}{(m_2+1)^2} dm_2 \\
= \;& \oint \left\{ \frac{\lambda}{m_2^2} - \frac{\lambda\gamma}{(m_2+1)^2} + \frac{1}{m_2} + \frac{1}{m_2+1} - \frac{1}{m_2-M} - \frac{1}{m_2-N} \right\} \frac{1}{m_2+1} dm_2 \\
= \;& 2\pi i \left(\lambda - 1 + \frac{1}{1+N} \right) . \qquad (27)
\end{aligned}$$

Applying Lemma 1 and Cauchy's formula, we obtain

$$\oint \frac{g(z(m_2))}{m_2+1} dm_2 = -2\pi i \left\{ \log(1-\lambda)(1+N) \right\} \qquad (28)$$

and

$$\oint \frac{g(z(m_2))}{m_2-M} dm_2 = 2\pi i \left\{ -\frac{\lambda}{M} - \lambda - \log \frac{(1-\lambda)(M-N)}{M} \right\} \qquad (29)$$

Combining the results (26) - (29), we have the desired result of variance. □

A.2 Proof of Theorem 2

We compute

$$\int g(x) dF^{\tilde{\gamma}', H_p}(x) = \int \left\{ \lambda x - \lambda - \log(\lambda x + 1 - \lambda) \right\} dF^{\gamma', H_p}(x),$$

where H_p is the SD of spiked population model, which is written as

$$H_p(t) = \frac{p-K}{p} \delta_1(t) + \frac{1}{p} \sum_{i=1}^{K} n_i \delta_{a_i}(t).$$

Following the lines of Section 3 in Wang et al. (2014),

$$\int \left(\lambda x - \lambda \right) dF^{\tilde{\gamma}', H_p}(x) = \frac{\lambda}{p} \sum_{i=1}^{K} n_i a_i - \frac{\lambda K}{p} + O(\frac{1}{n^2}).$$

The difficult part lies on the evaluation of integration of the logarithm-related term. Using the labeling in Proposition 2, we rewrite it as

$$\int \log(\lambda x + 1 - \lambda) dF^{\tilde{\gamma}', H_p}(x) = (14) + (15) + (16)$$

and calculate (14), (15) and (16) separately. In the remainder of the proof, we write m and γ instead of $\underline{m}$ and $\tilde{\gamma}'$, respectively, for notational convenience. The terms (14) and (15)

involve contour integrals. Recall that the contour $\mathcal{C}$ on (14) and (15) encloses the closed interval $[\frac{-1}{1-\sqrt{\gamma}}, \frac{-1}{1+\sqrt{\gamma}}]$ on the real axis of the complex plane and has poles of $\{m = -1\}$, $\{m = M\}$, where $M = M(\lambda, \gamma)$ is defined in (19). It is easy to show that $\frac{-1}{1-\sqrt{\gamma}} < M < \frac{-1}{1+\sqrt{\gamma}}$ and $N > 0$ provided $\gamma \in (0,1)$ and n is large, where $N = N(\lambda, \gamma)$ is from (19).

Following the lines of Section 3.3 of Wang et al. (2014), recall that $\lambda(-\frac{1}{m}+\frac{\gamma}{1+m})+(1-\lambda) = \frac{(1-\lambda)(m-M)(m-N)}{m(m+1)}$. We have

$$
\begin{aligned}
&(14)\\
= \; & \frac{-1}{2\pi i p}\oint_{\mathcal{C}} \log\Big(\lambda(-\frac{1}{m}+\frac{\gamma}{1+m})+(1-\lambda)\Big)\cdot\Big(\frac{K}{\gamma m}-\sum_{i=1}^{K}\frac{n_i a_i^2 m}{(1+a_i m)^2}\Big)dm\\
= \; & \frac{-1}{2\pi i p\gamma}\oint_{\mathcal{C}} \frac{\log\Big(\lambda(-\frac{1}{m}+\frac{\gamma}{1+m})+(1-\lambda)\Big)}{m}\cdot\Big(K-\sum_{i=1}^{K}\frac{n_i a_i^2 m^2\gamma}{(1+a_i m)^2}\Big)dm\\
= \; & \frac{-1}{2\pi i p\gamma}\oint_{\mathcal{C}} \frac{\log\big(\frac{(1-\lambda)(m-N)}{m}\big)+\log\big(\frac{m-M}{m+1}\big)}{m}\cdot\Big(K-\sum_{i=1}^{K}\frac{n_i a_i^2 m^2\gamma}{(1+a_i m)^2}\Big)dm\\
= \; & \frac{-K}{2\pi i p\gamma}\oint_{\mathcal{C}} \frac{\log\big(\frac{m-M}{m+1}\big)}{m}dm+\frac{1}{2\pi i p\gamma}\oint_{\mathcal{C}}\log\big(\frac{m-M}{m+1}\big)\sum_{i=1}^{K}\frac{n_i a_i^2 m\gamma}{(1+a_i m)^2}dm\\
\triangleq \; & A_1 + A_2.
\end{aligned}
$$

Here,

$$
\begin{aligned}
A_1 &= \frac{-K}{2\pi i p\gamma}\oint_{\mathcal{C}}\log\big(\frac{m-M}{m+1}\big)d\log m\\
&= \frac{K}{2\pi i p\gamma}\oint_{\mathcal{C}}\log m\cdot d\log\big(\frac{m-M}{m+1}\big)\\
&= \frac{K}{2\pi i p\gamma}\cdot(M+1)\cdot\oint_{\mathcal{C}}\frac{\log m}{(m+1)(m-M)}dm\\
&= \frac{K}{p\gamma}\log(-M),
\end{aligned}
$$

and

$$
\begin{aligned}
A_2 &= \frac{1}{2\pi i p}\oint_{\mathcal{C}}\log\big(\frac{m-M}{m+1}\big)\sum_{i=1}^{K}\frac{n_i a_i^2 m}{(1+a_i m)^2}dm\\
&= \frac{1}{2\pi i p}\sum_{i=1}^{K}\oint_{\mathcal{C}}\log\big(\frac{m-M}{m+1}\big)\cdot n_i a_i\Big(\frac{1}{1+a_i m}-\frac{1}{(1+a_i m)^2}\Big)dm\\
&\triangleq A_3 - A_4,
\end{aligned}
$$

where

$$\begin{aligned}
A_3 &= \frac{1}{2\pi ip}\sum_{i=1}^{K}\oint_{\mathcal{C}}\log\big(\frac{m-M}{m+1}\big)\frac{n_i a_i}{1+a_i m}dm \\
&= \frac{1}{2\pi ip}\sum_{i=1}^{K}\oint_{\mathcal{C}} n_i\log\big(\frac{m-M}{m+1}\big)d\log(1+a_i m) \\
&= \frac{-1}{2\pi ip}\sum_{i=1}^{K}\oint_{\mathcal{C}} n_i\log(1+a_i m)\cdot d\log\big(\frac{m-M}{m+1}\big) \\
&= \frac{-1}{2\pi ip}\cdot(M+1)\sum_{i=1}^{K}\oint_{\mathcal{C}}\frac{n_i\log(1+a_i m)}{(m+1)(m-M)}dm \\
&= \frac{1}{p}\sum_{i=1}^{K}n_i\log(1-a_i)-\frac{1}{p}\sum_{i=1}^{K}n_i\log(1+a_i M),
\end{aligned}$$

and

$$\begin{aligned}
A_4 &= \frac{1}{2\pi ip}\sum_{i=1}^{K}\oint_{\mathcal{C}}\log\big(\frac{m-M}{m+1}\big)\frac{n_i a_i}{(1+a_i m)^2}dm \\
&= \frac{1}{2\pi ip}\sum_{i=1}^{K}\oint_{\mathcal{C}}\frac{n_i}{1+a_i m}\cdot d\log\big(\frac{m-M}{m+1}\big) \\
&= \frac{M+1}{2\pi ip}\sum_{i=1}^{K}\oint_{\mathcal{C}}\frac{n_i}{(1+a_i m)(m-M)(m+1)}dm \\
&= \frac{1}{p}\sum_{i=1}^{K}n_i\big(\frac{1}{1+a_i M}-\frac{1}{1-a_i}\big).
\end{aligned}$$

Combiing the results of $A_1+A_2=A_1+A_3-A_4$, we have

$$(14)=\frac{K}{p\gamma}\log(-M)+\frac{1}{p}\sum_{i=1}^{K}n_i\log\Big(\frac{1-a_i}{1+a_i M}\Big)-\frac{1}{p}\sum_{i=1}^{K}n_i\Big(\frac{1}{1+a_i M}-\frac{1}{1-a_i}\Big).$$

Next, consider (15). Taking $f=\log\circ\psi$, we have

$$(\log\circ\psi)'\big(-\frac{1}{m}+\frac{\gamma}{1+m}\big)=\frac{\lambda}{\lambda x+(1-\lambda)}\bigg|_{x=-\frac{1}{m}+\frac{\gamma}{1+m}}=\frac{\lambda m(m+1)}{(1-\lambda)(m-M)(m-N)}.$$

Then

$$
\begin{aligned}
&(15)\\
= &\ -\frac{1}{2\pi ip}\oint_{\mathcal{C}}(\log\circ\psi)'\big(-\frac{1}{m}+\frac{\gamma}{1+m}\big)\sum_{i=1}^{K}\big(\frac{n_ia_i}{1+a_im}-\frac{n_i}{1+m}\big)\big(\frac{1}{m}-\frac{\gamma m}{(1+m)^2}\big)dm\\
= &\ -\frac{1}{2\pi ip}\cdot\frac{\lambda}{1-\lambda}\sum_{i=1}^{K}n_i\oint_{\mathcal{C}}\frac{m(m+1)}{(m-M)(m-N)}\big(\frac{a_i}{1+a_im}-\frac{1}{1+m}\big)\big(\frac{1}{m}-\frac{\gamma m}{(1+m)^2}\big)dm\\
= &\ -\frac{1}{2\pi ip}\cdot\frac{\lambda}{1-\lambda}\sum_{i=1}^{K}n_i(B_1-B_2-B_3+B_4),
\end{aligned}
$$

where

$$
\begin{aligned}
B_1 &= \oint_{\mathcal{C}}\frac{a_i(m+1)}{(m-M)(m-N)(1+a_im)}dm\\
&= 2\pi i\cdot\left(\frac{a_i(M+1)}{(1+a_iM)(M-N)}\right),\\
B_2 &= \oint_{\mathcal{C}}\frac{a_i\gamma m^2}{(m-M)(m-N)(1+a_im)(m+1)}dm\\
&= 2\pi i\cdot\left(\frac{a_i\gamma M^2}{(M-N)(1+a_iM)(M+1)}+\frac{a_i\gamma}{(M+1)(N+1)(1-a_i)}\right),\\
B_3 &= \oint_{\mathcal{C}}\frac{1}{(m-M)(m-N)}dm = 2\pi i\cdot\frac{1}{M-N},
\end{aligned}
$$

and

$$
\begin{aligned}
B_4 &= \oint_{\mathcal{C}}\frac{\gamma m^2}{(m-M)(m-N)(m+1)^2}dm\\
&= 2\pi i\cdot\left(\frac{\gamma M^2}{(M-N)(M+1)^2}-\frac{\gamma(2MN+M+N)}{(M+1)^2(N+1)^2}\right).
\end{aligned}
$$

Collecting the four terms, we have :

$$
\begin{aligned}
&(15)\\
= &\ -\frac{\lambda}{p(1-\lambda)}\cdot\sum_{i=1}^{K}n_i\Bigg[\frac{1}{M-N}\left\{\frac{a_i(M+1)}{1+a_iM}-\frac{a_i\gamma M^2}{(1+a_iM)(M+1)}-1+\frac{\gamma M^2}{(M+1)^2}\right\}\\
&\qquad -\frac{1}{(M+1)(N+1)}\left\{\frac{a_i\gamma}{1-a_i}+\frac{\gamma(2MN+M+N)}{(M+1)(N+1)}\right\}\Bigg].
\end{aligned}
$$

To obtain (16), note that the integration term $\int\log(\psi(x))dF^{\gamma,\delta_1}(x)$ is equal to $-\int\{\psi(x)+$

$\log(\psi(x)) - 1\}dF^{\gamma,\delta_1}(x)$ since M-P law satisfies $\int x dF^{\gamma,\delta_1}(x) = \int 1 dF^{\gamma,\delta_1}(x)$. This gives

$$(16) = -\left(1 - \frac{K}{p}\right) \int g(x) dF^{\gamma,\delta_1}(x) + \frac{1}{p}\sum_{i=1}^{K} n_i \log\psi\{\varphi(a_i)\} + O(\frac{1}{n^2})$$

where $\varphi(a_i) = a_i + \frac{\gamma a_i}{a_i - 1}$.

Finally, combining the four results, we finally obtain the centering term :

$$\begin{aligned} & \int \big\{\psi(x) - \log(\psi(x)) - 1\big\} dF^{\gamma,H_p}(x) \\ = & \int \big(\lambda x - \lambda\big) dF^{\gamma,\delta_1}(x) - (14) - (15) - (16). \\ = & \left(1 - \frac{K}{p}\right) \int g(x) dF^{\gamma,\delta_1}(x) + \frac{1}{p} C(\lambda, \gamma + O(\frac{1}{n^2}), \end{aligned}$$

where

$$\begin{aligned} C_n \;=\; & \lambda \cdot \frac{1}{K}\sum_{i=1}^{K} n_i a_i - \lambda - \frac{1}{K}\sum_{i=1}^{K} n_i \log\psi\{\varphi(a_i)\} \\ & - \left[\frac{1}{\gamma}\log(-M) + \frac{1}{K}\sum_{i=1}^{K} n_i \log\Big(\frac{1 - a_i}{1 + a_i M}\Big) - \frac{1}{K}\sum_{i=1}^{K} n_i \Big(\frac{1}{1 + a_i M} - \frac{1}{1 - a_i}\Big)\right] \\ & + \frac{\lambda}{(1-\lambda)K}\sum_{i=1}^{K} n_i \left[\frac{1}{M - N}\left\{\frac{a_i(M+1)}{1 + a_i M} - \frac{a_i \gamma M^2}{(1 + a_i M)(M+1)} - 1 + \frac{\gamma M^2}{(M+1)^2}\right\}\right. \\ & \left. - \frac{1}{(M+1)(N+1)}\left\{\frac{a_i \gamma}{1 - a_i} + \frac{\gamma(2MN + M + N)}{(M+1)(N+1)}\right\}\right]. \quad \square \end{aligned}$$

A.3 Technical lemma

Lemma 1. *Let z_A and z_B be any two different fixed complex numbers. Then, (a) for any contour $\mathcal{C}$ enclosing z_A and z_B such that*

$$\log\frac{z - z_A}{z - z_B}$$

is single-valued on the contour, we have

$$\oint \frac{1}{z - z_A}\log\frac{z - z_A}{z - z_B} dz = 0\,,$$

(b) for any contour $\mathcal{C}$ enclosing z_A and z_B , we have

$$\oint \frac{dz}{(z - z_A)(z - z_B)} = 0$$

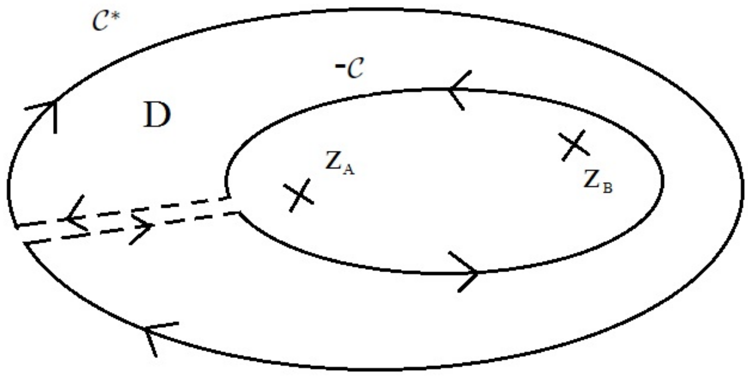

Figure 3: Illustration of the path of the contour integral

Proof. (a) Let $\mathcal{C}^*$ be a contour enclosing $\mathcal{C}$ such that both $\mathcal{C}$ and $\mathcal{C}^*$ are clockwise (anticlockwise), and on $\mathcal{C}^*$, we have $|z-z_A| < |z_B - z_A|$ (see Figure 3). Then, D , the singly connected region between $\mathcal{C}$ and $\mathcal{C}^*$ as indicated in Figure 3 contains no singularity. Therefore, the integral on $\mathcal{C}$ and $\mathcal{C}^*$ are the same. Next, consider the power series expansion

$$\begin{aligned}\log \frac{z-z_A}{z-z_B} &= -\log\left(1+\frac{z_A-z_B}{z-z_A}\right)\\ &= -\left\{\left(\frac{z_A-z_B}{z-z_A}\right)-\frac{1}{2}\left(\frac{z_A-z_B}{z-z_A}\right)^2+\frac{1}{3}\left(\frac{z_A-z_B}{z-z_A}\right)^3-\dots\right\}.\end{aligned}$$

Such power series converges on $\mathcal{C}^*$. The desired result is a consequence of

$$\oint \frac{dz}{(z-z_A)^i} = 0$$

for all $i = 2, 3, \dots$.

(b) Applying Cauchy's formula, we have

$$\oint \frac{dz}{(z-z_A)(z-z_B)} = \frac{1}{z_A-z_B}\oint\left\{\frac{1}{z-z_A}-\frac{1}{z-z_B}\right\}dz = \frac{1-1}{z_A-z_B} = 0.$$

□

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley-Interscience, 3rd edition.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics*, 37(6B):3822–3840.

Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177.

Bai, Z. D. and Silverstein, J. W. J. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605.

Cai, T. T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388.

Chen, L. S., Paul, D., Prentice, R. L., and Wang, P. (2011). A regularized Hotelling's T-test for pathway analysis in proteomic studies. *Journal of the American Statistical Association*, 106(496):1345–1360.

Chen, S., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Jiang, D., Jiang, T., and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference*, 142(8):2241–2256.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Pan, G. (2014). Comparison between two types of large sample covariance matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(2):655–677.

Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *The Annals of Applied Probability*, 18(3):1232–1270.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).

Wang, C., Yang, J., Miao, B., and Cao, L. (2013). Identity tests for high dimensional data using RMT. *Journal of Multivariate Analysis*, 118:128–137.

Wang, Q., Silverstein, J. W., and Yao, J.-f. (2014). A note on the CLT of the LSS for sample covariance matrix from a spiked population model. *Journal of Multivariate Analysis*, 130:194–207.

Warton, D. I. (2008). Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association*, 103(481):340–349.

Won, J.-H., Lim, J., Kim, S.-J., & Rajaratnam, B. (2013). Condition number regularized covariance estimation. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 75(3):427–450.

Zheng, S., Bai, Z. D, and Yao, J-F. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, 43:546-591.